Equalization in digital data Transmission systems

In the previous section we showed that by specifying $H_R(f)$ and $H_T(f)$ we can obtain zero ISI and minimum prob. of errors. In some instances it may be difficult to realize filters. In other cases, the channel may be unknown.

In all these cases, a filter with adjustable $P_R$ response would be useful to employ at the receiver. This filter is referred to as an equalization filter, or simply equalizer. We have two kinds of equalizers: one is called preset and the other is adaptive.
Zero-Forcing Equalizers (Preset):

Consider the following block diagram:

\[ H_0(f) = H_T(f) H_c(f) H_E(f) \]

For zero ISI we should have:

\[ \sum_{k=-\infty}^{\infty} H_0(f + \frac{k}{T}) = \text{Constant} \quad \text{if} \quad |f| \leq \frac{1}{2T} \]

\( h_E(t) \) is often approximated by a FIR filter as shown below:
In matrix form we have

\[ \mathbf{P}_{eq} = [\mathbf{P}_c] \mathbf{C} = \mathbf{D} \mathbf{C} = [\mathbf{P}_c]^{-1} \mathbf{P}_{eq} \]

or

\[
\begin{pmatrix}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{pmatrix}
= \begin{pmatrix}
P_c(0) & P_c(-1) & \cdots & P_c(-N) \\
& & \ddots & \vdots \\
& & & P_c(0)
\end{pmatrix}
\begin{pmatrix}
C_N \\
C_{-N+1} \\
\vdots \\
C_N
\end{pmatrix}
\]

\((2N+1) \times (2N+1)\)

Channel response

\(\mathbf{D}\) column 1

Notice that the sample values of the channel pulse response taken at \(T_{sec}\) is being used to evaluate \(2N+1\) unknown coefficients \(C_N, C_{-N+1}, \ldots, C_N\) for zero ISI.

So example 3-8 \(\frac{T_{sec}}{150}\)
the equalizer pulse response, \( p_{eq}(t) \), due to the channel output pulse response, \( p_c(t) \), is

\[
p_{eq}(t) = \sum_{n=-N}^{N} C_n p_c(t - nT)
\]

(3-158)

The zero- ISI condition (3-110b), when applied to (3-158), can hold for only \( 2N + 1 \) sample times since only \( 2N + 1 \) unknown constants are available for adjustment. Setting \( t = mT + \Delta t \), \( m = 0, \pm 1, \pm 2, \ldots, \pm N \), in (3-158), these conditions are

\[
p_{eq}(mT + \Delta t) = \sum_{n=-N}^{N} C_n p_c[(m - n)T + \Delta t]
\]

\[
= \begin{cases} 
1, & m = 0 \\
0, & m \neq 0
\end{cases}
\]

(3-159)

where \( t = \Delta t \) is the sampling time for which \( p_{eq}(t) \) is maximum. These equations can be written in matrix form as

\[
P_{eq} = [P_c] C
\]

(3-160)

where \( P_{eq} \) and \( C \) are vectors or column matrices given by

\[
P_{eq} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}
\]

\[
[N \text{ zeros}]
\]

\[
C = \begin{bmatrix} C_{-N} \\ C_{-N+1} \\ \vdots \\ C_0 \\ C_1 \\ \vdots \\ C_N \end{bmatrix}
\]

respectively, and \([P_c]\) is the \((2N + 1) \times (2N + 1)\) matrix of channel responses of the form

\[
[P_c] = \begin{bmatrix} p_c(0) & p_c(-1) & \cdots & p_c(-2N) \\ p_c(1) & p_c(0) & \cdots & p_c(-2N + 1) \\ p_c(2) & p_c(1) & \cdots & p_c(-2N + 2) \\ \vdots & \vdots & \ddots & \vdots \\ p_c(2N) & p_c(2N - 1) & \cdots & p_c(0) \end{bmatrix}
\]

(3-162)

The maximum of the channel pulse response, denoted by \( p_c(0) \), occurs at time \( t_m = \Delta t \). Thus the \( 4N + 1 \) sample values of the channel pulse response taken at \( T \)-second intervals can be used to determine the \( 2N + 1 \) unknown coefficients \( C_{-N}, C_{-N+1}, \ldots, C_0, \ldots, C_N \) by solving (3-162). Exactly \( N \) zeros will be forced at the sampling instants either side of the main pulse response. Because all components of the vector \( P_{eq} \) are zeros except for the center one, it follows that the coefficient vector \( C \) is the center column of \([P_c]^{-1}\).
The coefficient vector is the center column of \( [P_r]^{-1} \). Therefore,
\[
\begin{align*}
C_{-2} &= 0.117 & C_{-1} &= -0.158 & C_0 &= 0.937 \\
C_1 &= 0.133 & C_2 &= -0.091
\end{align*}
\]
The sample values of the equalized pulse response, from (3-159), are
\[
p_{eq}(m) = \sum_{n=-2}^{2} C_n p_r(m - n) = 1.0
\]  
where \( \Delta t = 0 \). For example,
\[
p_{eq}(0) = (0.117)(0.1) + (-0.158)(-0.1) + (0.937)(1) + (0.133)(0.2) + (-0.091)(-0.1) = 1.0
\]
which checks with the desired value of unity. Similarly, it can be verified that \( p_{eq}(-2) = p_{eq}(-1) = p_{eq}(1) = p_{eq}(2) = 0 \). Values of \( p_{eq}(n) \) for \( n < -2 \) or \( n > 2 \) are not zero. For example,
\[
p_{eq}(3) = (0.117)(0.005) + (-0.158)(0.02) + (0.937)(-0.05) + (0.133)(0.1) + (-0.091)(-0.1)
\]
\[
= 0.027
\]
\[
p_{eq}(-3) = (0.117)(0.2) + (-0.158)(-0.1) + (0.937)(-0.05) + (0.133)(-0.02) + (-0.091)(0.01)
\]
\[
= 0.082
\]
The zero-forcing equations (3-160) do not account for the effects of noise. In addition, the finite-length transversal filter equalizer can minimize worst-case ISI only if the peak distortion is less than 100% of the eye opening. Another type of equalizer which partially avoids these problems is the minimum-mean-square-error (MMSE) equalizer. In an MMSE equalizer, the equalizer coefficients are chosen to minimize the mean-square error, which consists of the sum of the squares of all the ISI terms plus the noise power at the equalizer output. The MMSE equalizer therefore maximizes the signal-to-distortion ratio at its output within the constraints of the equalizer length and delay.

### 3.9.2 Minimum Mean-Square Error Equalization

Suppose that the desired output from the transversal filter equalizer of Figure 3-35 is \( d(t) \). It is desired to choose the filter tap weights so that the mean-square error between desired output and its actual output is minimized. Since the actual output includes noise, we denote it as \( z(t) \) to distinguish it from the pulse response of the transversal filter. The MMSE criterion at the filter output can therefore be expressed as
\[
E = E[(z(t) - d(t))^2] = \text{minimum}
\]
For a transversal filter input denoted by \( y(t) \), which includes AWGN, the filter output is
\[
z(t) = \sum_{n=-N}^{N} C_n y(t - nT)
\]
Since the mean-square error is a concave function of the tap weights, a set of sufficient conditions for minimizing the mean-square error is
\[
\frac{\partial E}{\partial C_m} = 0 = 2E\left[ (z(t) - d(t)) \frac{\partial z(t)}{\partial C_m} \right], \quad m = 0, \pm 1, \pm 2, \ldots, \pm N
\]
Substitution of (3-168) into (3-169) and differentiation results in the set of conditions
\[
E[(z(t) - d(t)y(t - mT))] = 0, \quad m = 0, \pm 1, \pm 2, \ldots, \pm N
\]
or
\[
R_{yy}(mT) = R_{yd}(mT), \quad m = 0, \pm 1, \pm 2, \ldots, \pm N
\]
where
\[
R_{xy}(\tau) = E[y(t)z(t + \tau)]
\]
and
\[
R_{yd}(\tau) = E[y(t)d(t + \tau)]
\]
are the cross-correlation functions of the received signal with the transversal filter output and with the data, respectively.

If we substitute (3-168) for \( z(t) \) into (3-171), interchange the order of summation, and take the expectation, the conditions for the optimum MMSE transversal filter weights become the matrix equation
\[
[R_{yy}]C = [R_{yd}]
\]
where
\[
[R_{yy}] = \begin{bmatrix}
R_{yy}(0) & R_{yy}(T) & \cdots & R_{yy}(2NT) \\
R_{yy}(-T) & R_{yy}(0) & \cdots & R_{yy}(2(N-1)T) \\
\vdots & \vdots & \ddots & \vdots \\
R_{yy}(-2NT) & R_{yy}(-2(N-1)T) & \cdots & R_{yy}(0)
\end{bmatrix}
\]
and
\[
[R_{yd}] = \begin{bmatrix}
R_{yd}(-NT) \\
R_{yd}(-(N-1)T) \\
\vdots \\
R_{yd}(NT)
\end{bmatrix}
\]
A adaptive Tapped-dly line filter:

Consider a tapped-dly line filter that consists of delay elements, a set of multipliers, adjustable weights, and a summer.

\[ y_k = \sum_{i=0}^{N-1} w_i x_{k-i} \]

The criterion of interest is the mean-square-error defined by

\[ E = \sum_{k} e_k^2 \]
\[
\text{when } \quad e_k = y_k - d_k
\]

\[
\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \left( \sum_k e_k^2 \right) = 2 \sum_k e_k \frac{\partial e_k}{\partial w_i}
\]

\[
= 2 \sum_k e_k \frac{\partial (y_k - d_k)}{\partial w_i}
\]

\[
= 2 \sum_k e_k \frac{\partial y_k}{\partial w_i} = 2 \sum_k e_k x_{k-i}
\]

Note that the last summation is recognized as the (deterministic) cross-correlation of \( e_k \) and \( x_{k-i} \), that is,

\[
R_{ex}(iT) = \sum_k e_k x_{k-i}
\]

\[
\therefore \quad \frac{\partial E}{\partial w_i} = 2 R_{ex}(iT)
\]

The optimality condition for minimum error is

\[
\frac{\partial E}{\partial w_i} = 0 \quad \text{or} \quad R_{ex}(iT) = 0 \quad \text{for} \quad i = 0, 1, \ldots, N-1.
\]
this means, for min error, the cross-correlation between the output error sequence \( \{e_k\} \) and the input sequence \( \{x_k\} \) must have zeros for the \( N \) components. This is known as the principle of orthogonality.

Let us define the input auto-correlation as

\[
R_x(iT) = \sum_k x_k x_{k-i}
\]

and cross-correlation of \( d \) and \( x \) as

\[
R_{dx} = \sum_k d_k x_{k-i}
\]

Then,

\[
\frac{\text{COE}}{\text{COW}} = 2 \sum_k e_k x_{k-i} = 2 \sum_{n=0}^{N-1} w_n R_x(iT-nT)
\]

\[
-2 \sum_{n=0}^{N-1} \sum_{k} w_n x_{k-n} R_d(iT) = 0
\]

\( i = 0, 1, \ldots, N-1 \).
Using matrix notation, we get

\[ \nabla = 2 \mathbf{R}_x \mathbf{W} - 2 \mathbf{R}_{dx} \]

where

\[ \nabla = \begin{pmatrix} \rho_1 / \rho w_1 \\ \vdots \\ \rho_3 / \rho w_{N-1} \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} w_0 \\ \vdots \\ w_{N-1} \end{pmatrix} \]

\[ \mathbf{R}_x = \begin{bmatrix} \mathbf{R}_x(0) & \mathbf{R}_x(-T) & \cdots & \mathbf{R}_x(-NT+T) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_x(NT-T) & \cdots & \cdots & \mathbf{R}_x(0) \end{bmatrix} \]

\[ \mathbf{R}_{dx} = \begin{bmatrix} \mathbf{R}_{dx}(1) \\ \vdots \\ \mathbf{R}_{dx}(NT-T) \end{bmatrix} \]
\[ \nabla = 0 = 0 \quad \nabla R_x W - \nabla R_{dx} = 0. \]

\[ \therefore \quad W_{opt} = R_x^{-1} R_{dx}. \]

Recall that

\[ E = \sum \frac{e^2}{\kappa} = \sum [y_k^2 + d_k^2 - 2y_k d_k]. \]

\[ \sum_{n=0}^{N-1} w_n x_{k-n} \]

\[ = \sum \left( \sum_{n=0}^{N-1} w_n x_{k-n} \right)^2 + \sum y_k^2 - 2 \sum \left( \sum_{n=0}^{N-1} w_n x_{k-n} \right) d_k \]

\[ = \sum \sum_{n=0}^{N-1} w_n w_m R_x (nT - mT) - 2 \sum_{n=0}^{N-1} w_n R_{dx} (nT) + \sum d_k^2. \]

\[ \text{Or, in matrix form} \]

\[ E = W^T R_x W - 2W^T R_{dx} + \sum d_k^2. \quad (1) \]
\[ E_{\min} = \begin{bmatrix} W \end{bmatrix}^T R_x \begin{bmatrix} W \end{bmatrix} - 2 \begin{bmatrix} W \end{bmatrix}^T R_d x + \sum_{k} d_k \]  \hspace{1cm} (2) \]

Subtracting (2) - (1): \hspace{1cm} 0

\[ E - E_{\min} = \begin{bmatrix} W - W_{opt} \end{bmatrix}^T R_x \begin{bmatrix} W - W_{opt} \end{bmatrix} \]

\[ \therefore \quad E = E_{\min} + \begin{bmatrix} W - W_{opt} \end{bmatrix}^T R_x \begin{bmatrix} W - W_{opt} \end{bmatrix} \]

A practical method of finding \( W_{opt} \) is to use the steepest-descent algorithm:

\[
\begin{cases}
\begin{align*}
\begin{bmatrix} W \end{bmatrix}_{k+1} &= \begin{bmatrix} W \end{bmatrix}_k - \frac{\alpha}{2} \nabla_k \\
\end{align*}
\end{cases}
\]

where \( \alpha \) is the step size, \( k = 0, 1, 2, \ldots \), and \( \begin{bmatrix} W \end{bmatrix}_0 = 0 \).
Operation of Equalizer:

There are two modes of operating an adaptive equalizer, as shown below.

The training sequence may, for example, consist of the linear maximal-length or Pseudo-noise (PN) sequence described below.

The output sequence is periodic with a period defined by

\[ N = 2^m - 1 \]
\[ R_x(t) = \begin{cases} 
A^2 \left(1 - \frac{N+1}{N^T} \right), & \text{if } \frac{NT}{N+1} \leq T \\
- \frac{A^2}{N} & \text{for the remainder of the period.}
\end{cases} \]
We see that $R_{x}(n)$ is somewhat similar to that of a random binary sequence.
Degradations due to realization imperfections in digital modulation systems:

So far, we have considered the channel in its ideal case. Also, bandlimited channels have also been studied and degradation due to ISI has been investigated. Now, we study other sources of imperfections. To proceed, consider the following block diagram:

The above model is a coherent communication system showing various system departures from ideal model.
1. Phase and amplitude Imbalance

(a) BPSK:

**Phase:**

A BPSK signal can be written as

\[ y(t) = A \sin \left[ 2\pi f_0 t + (1-\alpha) \frac{\pi}{2} d(t) \right] \]

if \( \alpha = 0 \), the above equation represents an ideal BPSK waveform.

![Diagram of BPSK signal processing](image)

\[ y_{BPSK}(t) \rightarrow \text{LPF} \rightarrow y(t) \]

\[ 2\cos(2\pi f_0 t) \text{ Coherent detector} \]

\[ y_D(t) = L \left\{ 2A \sin \left[ 2\pi f_0 t + (1-\alpha) \frac{\pi}{2} d(t) \right] \cos 2\pi f_0 t \right\} \]

\[ = A \sin \left[ (1-\alpha) \frac{\pi}{2} d(t) \right] \]

\[ = \begin{cases} 
 2A \sin \left( \frac{\pi}{2} - \frac{\pi \alpha}{2} \right) / \cos \left( \frac{\pi \alpha}{2} \right) & d(t) = 1 \\
 2A \sin \left( \frac{\pi}{2} + \frac{\pi \alpha}{2} \right) / \cos \left( \frac{\pi \alpha}{2} \right) & d(t) = -1 
\end{cases} \]

\[ y_D(t) = A d(t) \cos \frac{\pi \alpha}{2}. \]
notice that if \( \alpha = 0 \) (ideal BPSK) then
\[
\mathcal{Y}_d(t) = A d(t)
\]
The degradation, therefore, is:
\[
\frac{D}{P_s} = -20 \log_{10} \left( \frac{C_0 \frac{N d}{2}}{10} \right)
\]
notice that for \( \alpha = 0.5 \) we get
\[
\frac{D}{P_s} = -20 \log_{10} \left( \frac{C_0 \frac{\pi}{4}}{10} \right) \approx 3 \text{ dB}
\]

Amplitude:

An ideal BPSK waveform can be written as
\[
\mathcal{Y}_{BPSK}(t) = A \sin (2 \pi f_0 t + \pi) = \left\{ \begin{array}{ll}
A \cos 2 \pi f_0 t & \text{logic 1} \\
-A \cos 2 \pi f_0 t & \text{logic 0}
\end{array} \right.
\]
now assume that \( +1 \) corresponds to \( \pm 1 + \varepsilon \). then
\[
\mathcal{Y}_{BPSK}(t) = \left\{ \begin{array}{ll}
A (1+\varepsilon) \cos 2 \pi f_0 t & \text{logic 1} \\
A (-1+\varepsilon) \cos 2 \pi f_0 t & \text{logic 0}
\end{array} \right. = A \left[ d(t)+\varepsilon \right] \cos 2 \pi f_0 t.
\[
\frac{y_{D}(t)}{y_{BPSK}(t)} = \frac{2 \cos(2nf_0t)}{1 + \cos(4nf_0t)}
\]

Coherent detector:

\[
y_{D}(t) = L_{P}\left\{ 2A(\epsilon + d(t)) \cos^{2}(2nf_0t) \right\} = A(\epsilon + d(t))
\]

\[
\frac{\mu_{D}^{2}(t)}{\mu_{BPSK}^{2}} = A^{2}(\epsilon + d(t))^{2} = A^{2}(\epsilon^{2} + \overrightarrow{v(t)}^{2} + 2\epsilon \overrightarrow{v(t)}) = A^{2}(1 + \epsilon^{2})
\]

\[
D_{ai} = 10 \log_{10}(1 + \epsilon^{2})
\]
(b) **QPSK**:

**Phase**: Recall that an ideal QPSK is represented by

\[ x(t) = A \left[ d_1(t) \cos 2\pi f_c t + d_2(t) \sin 2\pi f_c t \right]. \]

Now, assume that there is an unbalance \( \beta \) between the two carriers.

\[ x_c(t) = A \left[ d_1(t) \cos (2\pi f_c t + \beta) + d_2(t) \sin (2\pi f_c t - \beta) \right]. \]

[Diagram showing the QPSK system with coherent detectors]
\[ y'_D(t) = L_p \left\{ x_c(t) - 2 \cos 2\pi f_0 t \right\} \]

\[ = L_p \left\{ 2A \frac{d_1(t)}{2} \cos (2\pi f_0 t + \frac{\beta}{2}) - \cos 2\pi f_0 t + 2A \frac{d_2(t)}{2} \frac{\sin (2\pi f_0 t - \frac{\beta}{2})}{\sin (2\pi f_0 t + \frac{\beta}{2})} \right\} \]

\[ = A \left[ \frac{d_1(t)}{2} \cos \frac{\beta}{2} - \frac{d_2(t)}{2} \frac{\sin \frac{\beta}{2}}{\sin \frac{\beta}{2}} \right] \]

Similarly,

\[ y'_D(t) = L_p \left\{ x_c(t) - 2 \sin 2\pi f_0 t \right\} \]

\[ = L_p \left\{ - \frac{d_1(t)}{2} \sin \frac{\beta}{2} + \frac{d_2(t)}{2} \cos \frac{\beta}{2} \right\} \]

\[ = A \left[ - \frac{d_1(t)}{2} \sin \frac{\beta}{2} + \frac{d_2(t)}{2} \cos \frac{\beta}{2} \right] \]

Notice that if \( \beta = 0 \) (ideal case) then

\[ \begin{align*}
    y'_D(t) &= A d_1(t) \\
    y'_D(t) &= A d_2(t)
\end{align*} \]
Phase imbalance introduces both amplitude degradation as well as crosstalk from the other channel.

Recall that the prob of error per channel is

\[ P_e = Q \left( \frac{\sqrt{2} E_b}{N_0} \right) \quad \text{ideal} \]

In phase imbalance situation we have

\[ y_D'(t) = A \left( \pm 1 \cos \frac{\beta}{2} \pm \sin \frac{\beta}{2} \right) \]

or

\[ y''_D(t) = A \left( \pm \sin \frac{\beta}{2} \pm \cos \frac{\beta}{2} \right) \]

\[ E_b' = A_{E_b} \left( \cos \frac{\beta}{2} + \sin \frac{\beta}{2} \right) \]

or

\[ E_b'' = A_{E_b} \left( \cos \frac{\beta}{2} - \sin \frac{\beta}{2} \right) \]

\[ P_e = \frac{1}{2} Q \left[ \sqrt{\frac{2 E_b}{N_0}} \left( \cos \frac{\beta}{2} + \sin \frac{\beta}{2} \right) \right] + \frac{1}{2} Q \left[ \sqrt{\frac{2 E_b}{N_0}} \left( \cos \frac{\beta}{2} - \sin \frac{\beta}{2} \right) \right] \]

Assuming that \( E_b' \) and \( E_b'' \) are equally likely. \( -\sin \left( \frac{\beta}{2} \right) \)
Amplitude:

\[ x_\psi(t) = A_1 d_1(t) \cos 2\pi f_0 t + A_2 d_2(t) \sin 2\pi f_0 t \]

\[ \theta(t) = -\frac{A_2 d'_2(t)}{A_1 d_1(t)} \]

The result is:

\[ y_1'(t) = A_1 d_1(t) \quad \text{no attenuation in demod output} \]

\[ y_2'(t) = A_2 d_2(t) \]

no cross-talk in this case!
Coherent Demodulator
(Imperfect Phase Reference):

Consider a BPSK waveform

\[ y_{BPSK}(t) \xrightarrow{\times} \text{LPF} \xrightarrow{\cos(2\pi f_0 t + \phi)} y_D(t) \]

Phase irregularity

Know:

\[ y(t) = A d(t) \cos(2\pi f_0 t) \]

\[ y_D(t) = \frac{1}{P} \left[ A d(t) \cos(2\pi f_0 t) \cos(2\pi f_0 t + \phi) \right]^2 \]

\[ = A d(t) \cos \phi. \]

\[ y_D^2(t) = \frac{A^2 d(t)^2 \cos^2 \phi}{1} = A^2 \cos^2 \phi \quad \text{dB degradation due to} \]

\[ D_{\text{BPSK}} = 10 \log A^2 \cos^2 \phi = 20 \log (\cos \phi) \quad \text{dB}. \]